Appendix

Appendix A1 Study characteristics: Schetz, 1994 (randomized controlled trial)

Characteristic	Description
Study citation	Schetz, K. F. (1994). An examination of software used with enhancement for preschool discourse skill improvement. *Journal of Educational Computing Research, 11*(1), 51–71. *Additional source:* Schetz, K. F. (1992). Preschool discourse skill improvement with computer-assisted instruction. *Dissertation Abstracts International, 52*(11), 3821-3822A. (UMI No. 9208433).
Participants	The study began with 97 low-income four- to five-year-old children. Four children were excluded from the study before assignment because of absences, and an additional 15 children were lost to attrition after assignment, leaving a sample of 78 children. The original sample of 93 children had a mean age of 4.6 years, 55% were female, and 17% were receiving speech-language services. The children were grouped into triads based on their summed pretest score, and one child from each triad was randomly assigned to the intervention and comparison conditions.
Setting	The study took place in five Head Start classes from three Head Start centers in the Roanoke Valley in Roanoke, Virginia.
Intervention	The study included two intervention groups: software with programmed instructional interaction with a speech-language clinician (or software with enhancement) and software without the programmed instructional component (or software without enhancement). The software used was from the *Words and Concepts* series (*Words and Concepts I, II, and III*), designed to teach children about nouns and concepts. In the software with enhancement condition, the children and the clinician interacted with the computer (that is, the clinician asked questions related to discourse skills and encouraged verbalization of responses to any questions from the clinician or the computer) to improve discourse skills (that is, receptive and expressive language). In the software without enhancement condition, the clinician gave the children instructions about how to use the software program to enhance their receptive vocabulary and concepts. In both conditions, children primarily participated in pairs twice weekly for 20 minutes a session over a period of 12 weeks.[1] The two groups were combined by the WWC for this review to determine the overall rating of effectiveness.[2] However, the WWC reports findings for the two intervention groups versus the comparison group separately in Appendix A4 and reports findings for the comparison between the two intervention groups in Appendices A5.1 and A5.2.
Comparison	Children in the no-treatment comparison group received language enrichment through the regular Head Start curriculum, including classroom activities (for example, house-keeping, circle time, dramatic play, finger plays, story time, and songs) and characteristics of the classroom environment (for example, labeling of classroom items). The same comparison group was used for both the software with enhancement and software without enhancement conditions.
Primary outcomes and measurement	The primary outcome domain was children's oral language, which was measured by two standardized tests: the Preschool Language Assessment Instrument (PLAI-I, PLAI-II, PLAI-III, PLAI-IV, and PLAI-Composite)[3] and the Peabody Picture Vocabulary Test-Revised (PPVT-R). (See Appendix A2 for more detailed descriptions of outcome measures.)[4]
Teacher training	The intervention was implemented by five student speech-language clinicians (two undergraduate students and three graduate students), who were trained and supervised by the researcher. Each classroom in the study had access to a speech-language clinician.

1. The author reported a number of barriers to implementation, including noisy and crowded facilities, faulty equipment, teacher turnover, and multiple sites to monitor.
2. The WWC recognizes that this is a different use of the data than intended by the study author, but combining the groups better addresses intervention effectiveness, which is a main task for the WWC. That is, the WWC is more concerned about the effects of *Words and Concepts* versus the comparison group than the possible effects of enhancement used in conjunction with *Words and Concepts.*
3. Schetz (1994) reported results for the PLAI-Composite and the PLAI-I, PLAI-II, PLAI-III, and PLAI-IV. For purposes of this report, the WWC includes the PLAI-Composite in the intervention ratings for the *Words and Concepts* series. The results for the PLAI-I, PLAI-II, PLAI-III, and PLAI-IV are included in the appendices, but do not contribute to the intervention rating.

(continued)

4. One child from each class from each intervention and comparison group was also randomly assigned to a language sample analysis (that is, three children from each class—one for each condition—for a total of 15 children). The WWC ECE team does not include the results of this analysis because of severe attrition within the subsample (posttest data were only available for 10 of the children) and incomplete reporting. Schetz (1994) also conducted two additional analyses (one comparing effects on high- and low-functioning children, the other estimating effects after excluding children receiving speech services). The author tested the interaction between treatment and level of functioning and found that there was no significant interaction. So, detailed findings for the lower and higher functioning subgroups are not included in the technical appendices. However, results for the PLAI-I suggest that *Words and Concepts* may be particularly useful for low-functioning children. The author also reestimated intervention effects excluding children receiving speech services and found that the results are not sensitive to the inclusion of children receiving speech services (that is, these analyses also showed no significant effects of the intervention).

Appendix A2 Outcome measures in the oral language domain

Outcome measure	Description
Peabody Picture Vocabulary Test-Revised (PPVT-R)	A standardized measure of children's receptive vocabulary that requires them to identify pictures that correspond to spoken words (as cited in Schetz, 1994).
Preschool Language Assessment Instrument (PLAI)	A standardized measure of children's communicative competence and perceptual language skills. In addition to the PLAI-Composite score, Schetz (1994) also used the following subscales: PLAI-I (to assess matching perception), PLAI-II (to assess selective analysis of perception), PLAI-III (to assess reordering perception), and PLAI-IV (to assess reasoning about perception). Questions from all four levels are placed at random throughout the test (as cited in Schetz, 1994).

Appendix A3 Summary of study findings included in the rating for the oral language domain[1]

Outcome measure	Study sample	Sample size (children)	Author's findings from the study: Mean outcome (standard deviation[2]) — *Words and Concepts* group[3]	Author's findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Words and Concepts* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Schetz, 1994 (randomized controlled trial; combined groups)[8]								
PLAI-Composite	4 year olds	78	101.74 (24.84)	96.75 (25.47)	4.99	0.20	ns	+8
PPVT-R	4 year olds	78	41.47 (13.73)	41.13 (16.02)	0.34	0.02	ns	+1
Domain average[9] for oral language						0.11	ns	+4

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Subscale and subgroup findings from the same study are not included in these ratings, but are reported in Appendix A4.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The intervention group mean equals the comparison group mean plus the mean difference.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean differences were computed by the WWC and took into account pretest difference between the study groups. The resulting effect sizes may overestimate the intervention's effects when the intervention group had lower pretest scores than the comparison group and underestimate the intervention's effects when the intervention group had higher pretest scores than the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study author or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Schetz (1994), no corrections for clustering or multiple comparisons were needed.
9. This row provides the study average, which in this case is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4 Summary of subscale and individual intervention group findings for the oral language domain[1]

Outcome measure	Study sample	Sample size (children)	Author's findings from the study: Mean outcome (standard deviation[2]) — *Words and Concepts* group[3]	Author's findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Words and Concepts* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Schetz, 1994 (randomized controlled trial; combined groups)[8]								
PLAI-I	4 year olds	78	37.12 (4.56)	35.24 (4.90)	1.88	0.40	ns	+15
PLAI-II	4 year olds	78	26.87 (7.46)	27.62 (8.35)	–0.75	–0.09	ns	–4
PLAI-III	4 year olds	78	22.93 (8.33)	20.93 (10.03)	2.00	0.22	ns	+9
PLAI-IV	4 year olds	78	14.81 (9.00)	12.96 (6.76)	1.85	0.22	ns	+9
Schetz, 1994 (randomized controlled trial; enhancement group)[8]								
PLAI-Composite	4 year olds	56	99.37 (26.04)	96.75 (25.47)	2.62	0.10	ns	+4
PPVT-R	4 year olds	56	41.60 (14.00)	41.13 (16.02)	0.47	0.03	ns	+1
PLAI-I	4 year olds	56	36.94 (5.17)	35.24 (4.90)	1.70	0.33	ns	+13
PLAI-II	4 year olds	56	26.43 (7.47)	27.62 (8.35)	–1.19	–0.15	ns	–6
PLAI-III	4 year olds	56	21.74 (8.44)	20.93 (10.03)	0.81	0.09	ns	+3
PLAI-IV	4 year olds	56	14.26 (8.73)	12.96 (6.76)	1.30	0.16	ns	+7
Schetz, 1994 (randomized controlled trial; without enhancement group)[8]								

(continued)

Appendix A4 Summary of subscale and individual intervention group findings for the oral language domain[1] *(continued)*

Outcome measure	Study sample	Sample size (children)	Author's findings from the study: Mean outcome (standard deviation[2]) — *Words and Concepts* group[3]	Author's findings from the study: Mean outcome (standard deviation[2]) — Comparison group[3]	WWC calculations: Mean difference[4] (*Words and Concepts* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
PLAI-Composite	4 year olds	51	104.65 (23.88)	96.75 (25.47)	7.90	0.31	ns	+12
PPVT-R	4 year olds	51	41.30 (13.69)	41.13 (16.02)	0.17	0.01	ns	0
PLAI-I	4 year olds	51	37.35 (3.72)	35.24 (4.90)	2.11	0.47	ns	+18
PLAI-II	4 year olds	51	27.42 (7.63)	27.62 (8.35)	–0.20	–0.02	ns	–1
PLAI-III	4 year olds	51	24.40 (8.37)	20.93 (10.03)	3.47	0.37	ns	+14
PLAI-IV	4 year olds	51	15.48 (9.52)	12.96 (6.76)	2.52	0.31	ns	+12

ns = not statistically significant

1. This appendix presents subscale and individual intervention group findings for measures that fall in the oral language domain. Total scale scores and combined groups were used for rating purposes and are presented in Appendix A3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The intervention group mean equals the comparison group mean plus the mean difference.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The mean differences were computed by the WWC and took into account pretest differences between the study groups. The resulting effect sizes may overestimate the intervention's effects when the intervention group had lower pretest scores than the comparison group and underestimate the intervention's effects when the intervention group had higher pretest scores than the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
8. The level of statistical significance was reported by the study author or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Schetz (1994), no correction for clustering was needed.

Appendix A5.1 Summary of findings for comparisons between *Words and Concepts* with enhancement and *Words and Concepts* without enhancement for the oral language domain[1]

Outcome measure	Study sample	Sample size (children)	Author's findings from the study: Mean outcome (standard deviation[2]) — *Words and Concepts* with enhancement group[3]	Author's findings from the study: Mean outcome (standard deviation[2]) — *Words and Concepts* without enhancement group[3]	WWC calculations: Mean difference[4] (*Words and Concepts* with enhancement – *Words and Concepts* without enhancement)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Schetz, 1994 (randomized controlled trial)[8]								
PLAI-Composite	4 year olds	49	96.44 (26.04)	101.72 (23.88)	–5.28	–0.21	ns	–8
PPVT-R	4 year olds	49	41.93 (14.00)	41.63 (13.69)	0.30	0.02	ns	+1
Domain average[9] for oral language						–0.09	ns	–4

ns = not statistically significant

1. This appendix presents a summary of study findings for measures that fall in the oral language domain for a comparison of two modes of implementation of *Words and Concepts* that is not included in the overall effectiveness rating.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The *Words and Concepts* with enhancement group mean equals the *Words and Concepts* without enhancement group mean plus the mean difference.
4. Positive differences and effect sizes favor the enhancement group; negative differences and effect sizes favor the without enhancement group. The mean differences were computed by the WWC and took into account pretest differences between the study groups. The resulting effect sizes may overestimate the enhancement groups' effects when the enhancement group had lower pretest scores than the without enhancement group and underestimate the enhancement groups' effects when the enhancement group had higher pretest scores than the without enhancement group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the enhancement condition and that of the average student in the without enhancement condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the enhancement group.
8. The level of statistical significance was reported by the study author or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Schetz (1994), no corrections for clustering or multiple comparisons were needed.
9. This row provides the study average, which in this case is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A5.2 Summary of subscale findings for comparisons between *Words and Concepts* with enhancement and *Words and Concepts* without enhancement for the oral language domain[1]

Outcome measure	Study sample	Sample size (children)	Author's findings from the study: Mean outcome (standard deviation[2]) — *Words and Concepts* with enhancement group[3]	Author's findings from the study: Mean outcome (standard deviation[2]) — *Words and Concepts* without enhancement group[3]	WWC calculations: Mean difference[4] (*Words and Concepts* with enhancement – *Words and Concepts* without enhancement)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Schetz, 1994 (randomized controlled trial)[8]								
PLAI-I	4 year olds	49	37.18 (5.17)	37.59 (3.72)	–0.41	–0.09	ns	–4
PLAI-II	4 year olds	49	27.19 (7.47)	28.18 (7.63)	–0.99	–0.13	ns	–4
PLAI-III	4 year olds	49	19.02 (8.44)	21.68 (8.37)	–2.66	–0.31	ns	–12
PLAI-IV	4 year olds	49	13.05 (8.73)	14.27 (9.52)	–1.22	–0.13	ns	–5

ns = not statistically significant

1. This appendix presents a summary of subscale findings for measures that fall in the oral language domain for a comparison that is not included in the overall effectiveness rating.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The *Words and Concepts* with enhancement group mean equals the *Words and Concepts* without enhancement group mean plus the mean difference.
4. Positive differences and effect sizes favor the enhancement group; negative differences and effect sizes favor the without enhancement group. The mean differences were computed by the WWC and took into account pretest differences between the study groups. The resulting effect sizes may overestimate the enhancement groups' effects when the enhancement group had lower pretest scores than the without enhancement group and underestimate the enhancement groups' effects when the enhancement group had higher pretest scores than the without enhancement group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the enhancement condition and that of the average student in the without enhancement condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the enhancement group.
8. The level of statistical significance was reported by the study author or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Schetz (1994), no corrections for clustering or multiple comparisons were needed.

Appendix A6 *Words and Concepts* rating for the oral language domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of oral language, the WWC rated *Words and Concepts* as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects because no studies showed statistically significant or substantively important effects, either positive or negative.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative.*

  Met. The study did not show statistically significant or substantively important effects, either positive or negative.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. Only one study examined the effects of *Words and Concepts* in the oral language domain.

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. The study did not show statistically significant or substantively important negative effects.

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Not met. The study did not show statistically significant or substantively important positive effects.

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Met. The one study that met WWC evidence standards did not show negative effects.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

  Not met. The study did not show statistically significant or substantively important effects, either positive or negative.

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

  Not met. The study did not show statistically significant or substantively important effects.

Potentially negative effects: Evidence of a negative effect with no overriding contrary evidence

(continued)

Appendix A6 *Words and Concepts* rating for the oral language domain *(continued)*

- Criterion 1: At least one study showing a statistically significant or substantively important *negative* effect.

  Not met. The study did not show statistically significant or substantively important negative effects.

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

  Met. The study did not show statistically significant or substantively important effects, either positive or negative.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *negative* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. Only one study examined the effects of *Words and Concepts* in the oral language domain.

- Criterion 2: No studies showing statistically significant or substantively important *positive* effects.

  Met. The study did not show statistically significant or substantively important positive effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effects. The WWC also considers the size of the domain level effects for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.